HBM IP Subsystem Implementation: 2.5D ASIC

Open-Silicon

TSMC 2016
Open Innovation Platform®
Ecosystem Forum

ABSTRACT

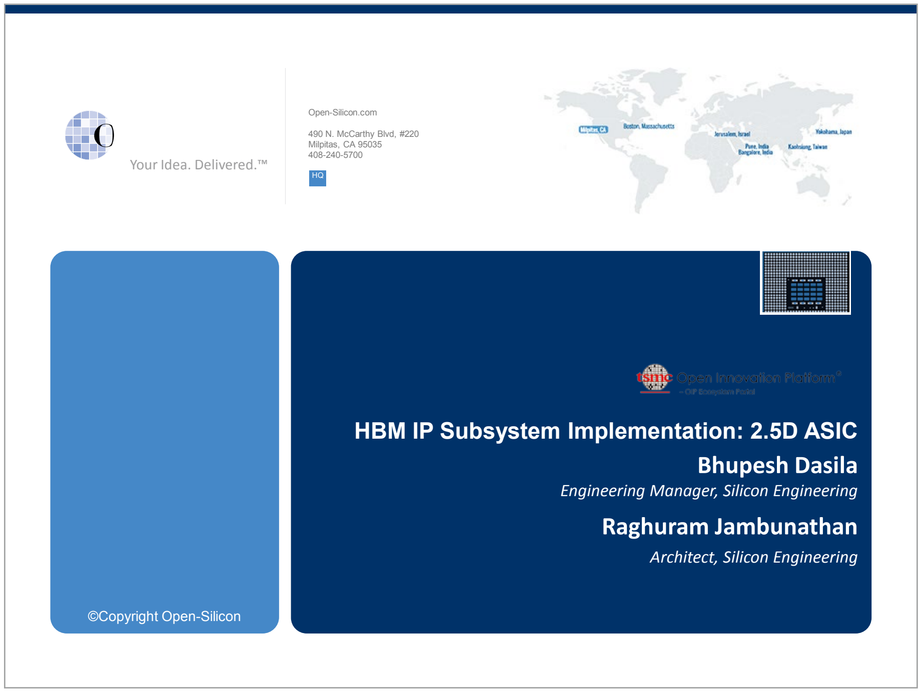
Access to memory is a critical aspect of successful high-end computing. The conventional approach to memory, where the growing proportion of the total power budget is consumed by memory, has become a limiting factor in overall system performance. Over the last decade, however, high memory bandwidth (HBM) at low power has increased significantly in various high-end computing applications such as graphics. This technology has overcome the prior limitations imposed by conventional memory.

HBM is a technology that ups the bandwidth-per-watt and the bandwidth-per-square-millimeter of die space. It provides high density, low power, low latency, fast flight time and high bandwidth at a lower pin count. HBM gen2 offers 256 GBytePS bandwidth, 8 Gb per die, 2Gbps per pin data rate, and only 2.5-3mw of power consumption per Gbps.

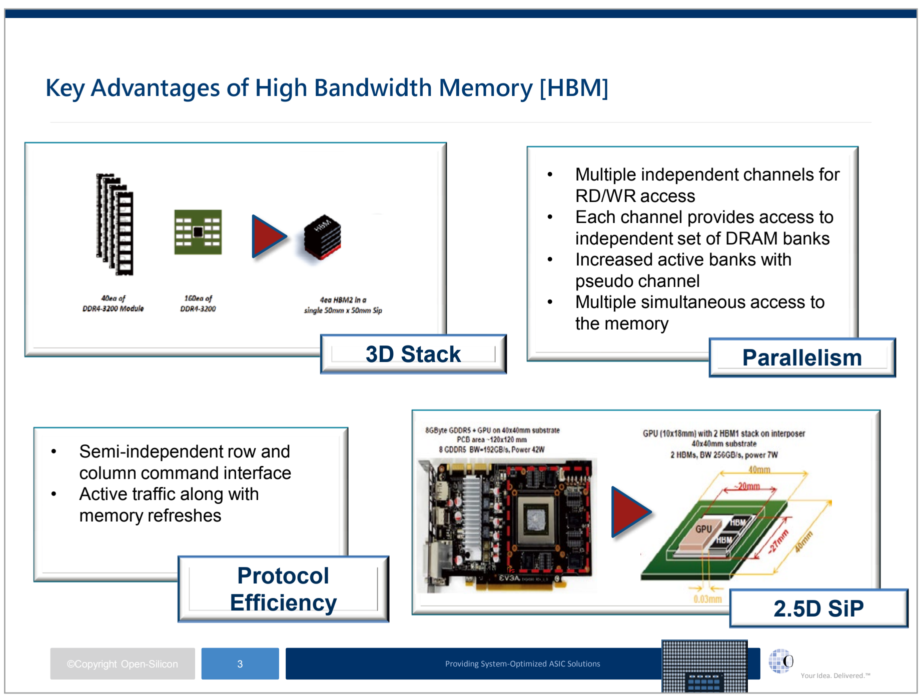
HBM ASIC System-in-Package (SiP) is a fantastic idea, but it presents various challenges with interposer design, die2die IOs, high speed interface, electrical tolerance, assembly of die, mechanical tolerances, ESD, packaging, test and so forth. Hence, the adoption of the 2.5D SiP technology has been slow. During the past several years, 2.5D silicon-interposer based SiP technologies have been proposed and evaluated for ASIC use. Thanks to the advent of JEDEC's, the HBM Gen-2 DRAM standard has advanced significantly. The assembly and manufacturing houses are now addressing many challenges associated with HBM 2.5D SiP. This includes things like the convergence of market needs, a higher memory bandwidth at lower power, and technical capabilities (2.5D Interposer-based ASIC design in combination with the new JEDEC HBM Gen2 standard). It needs the integration of significant capacities of high-bandwidth (up-to 256GB/s for an 8-channel, 8Gb memory stack implementation) and low latency memory inside the ASIC package.

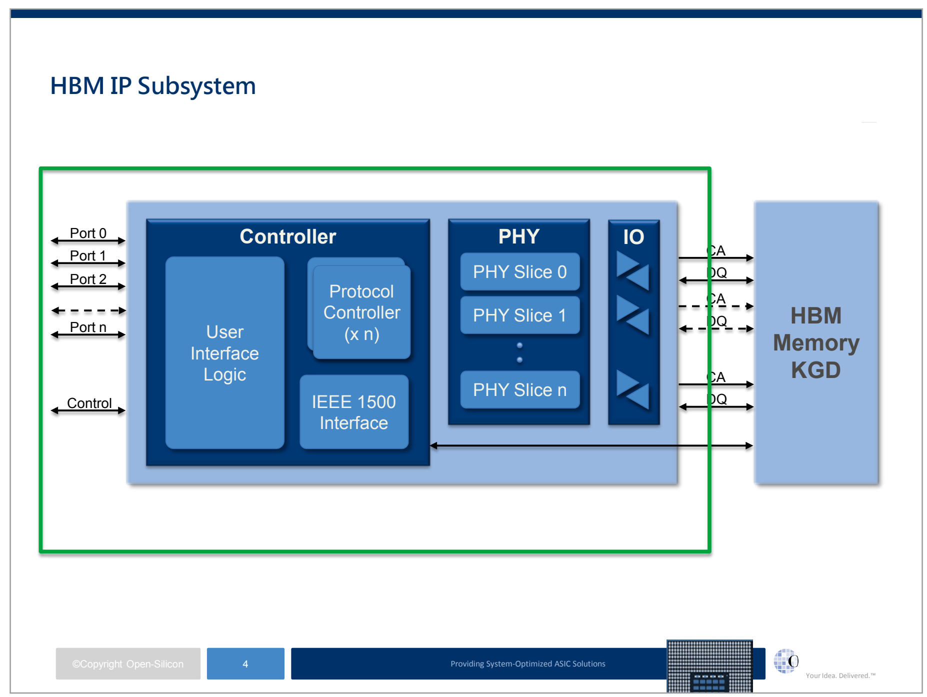
Considering the ever increasing requirement of low power, higher performance and low area, the 16nm FinFet process is the key to HBM subsystem 2.5D ASIC implementation. A 16nm FinFET process can potentially offer a 40-50% performance increase or a 50% power reduction compared to the 28nm process. The complete ecosystem (die, interposer, assembly, packaging) provided by manufacturing and assembly house is very crucial for adaptation of this technology.

During this presentation, Open-Silicon will present its HBM-Gen2 IP in TSMC's 16nm GL "FinFet" process in combination with TSMC's CoWoSTM 2.5D silicon-interposer technology. Aspects to be discussed include the HBM-controller, HBM-PHY and HBM die-to-die interposer-IO implementation, as well as the integration validation of the IP with SK- Hynix's HBM die-stack into a 16nm/2.5D/CoWoSTM ASIC implementation.

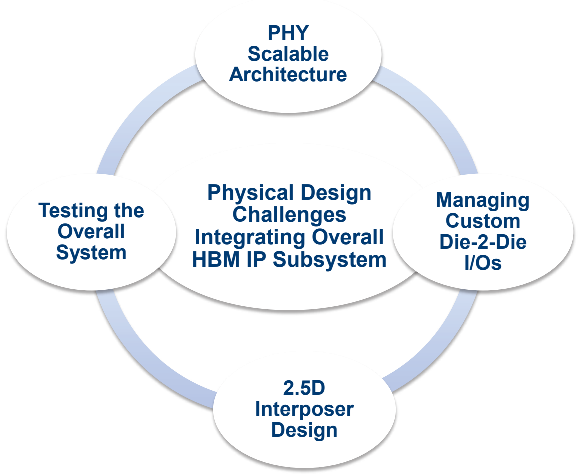

Your Idea. Delivered.™
Open-Silicon.com
490 N. McCarthy Blvd, #220
Milpitas, CA 95035
408-240-5700
HQ
HBM IP Subsystem Implementation: 2.5D ASIC
Bhupesh Dasila
Engineering Manager, Silicon Engineering
Raghuram Jambunathan
Architect, Silicon Engineering
©Copyright Open-Silicon


Requirements From Key Applications
Form Factor, Bandwidth, Power
Data Center
Networking
Radar
Memory Requirement: High bandwidth and density based on real-time-random operations
Virtual Reality
Gaming
Cloud Computing
©Copyright Open-Silicon
2
Providing System-Optimized ASIC Solutions

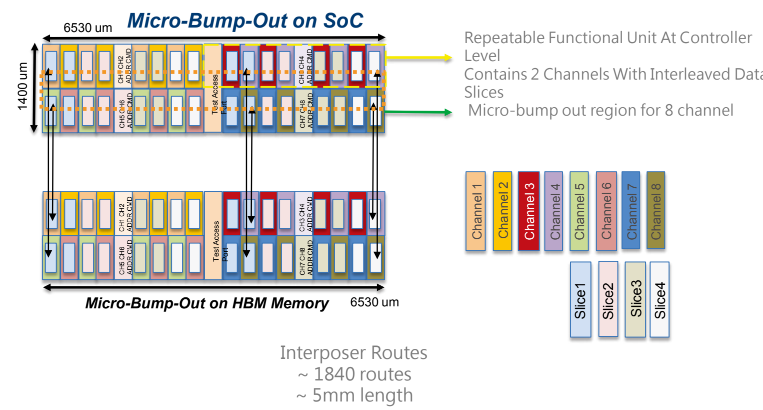

Key Advantages of High Bandwidth Memory [HBM]
40ea of DDR4-3200 Module
160ea of DDR4-3200
4ea HBM2 in a single 50mm x 50mm Sip
3D Stack
Multiple independent channels for RD/WR access
Each channel provides access to independent set of DRAM banks
Increased active banks with pseudo channel
Multiple simultaneous access to the memory
Parallelism
Semi-independent row and column command interface
Active traffic along with memory refreshes
Protocol Efficiency
2.5D SiP
©Copyright Open-Silicon
3
Providing System-Optimized ASIC Solutions


HBM IP Subsystem
Port 0
Port 1
Port 2
Port n
Control
Controller
User Interface Logic
Protocol Controller (x n)
IEEE 1500 Interface
PHY
PHY Slice 0
PHY Slice 1
PHY Slice n
IO
CA
DQ
CA
DQ
CA
DQ
HBM Memory KGD
©Copyright Open-Silicon
4
Providing System-Optimized ASIC Solutions

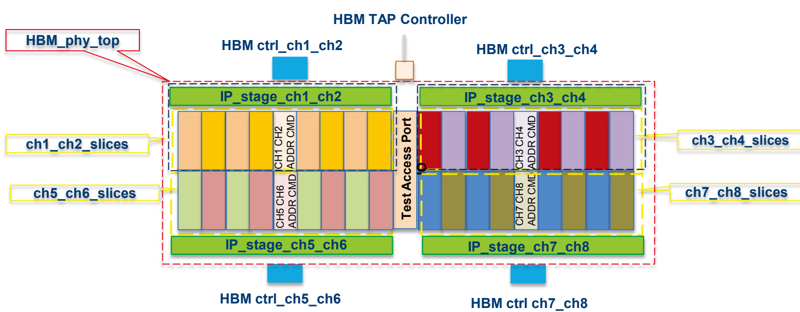
HBM 2.5D SiP Implementation Challenges



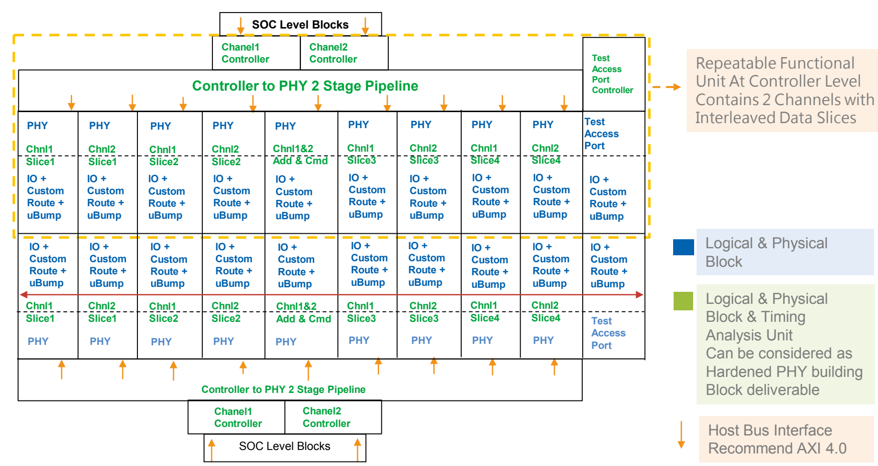
HBM Memory To Soc Micro-bump-out Interface



HBM PHY: Physical Hierarchy



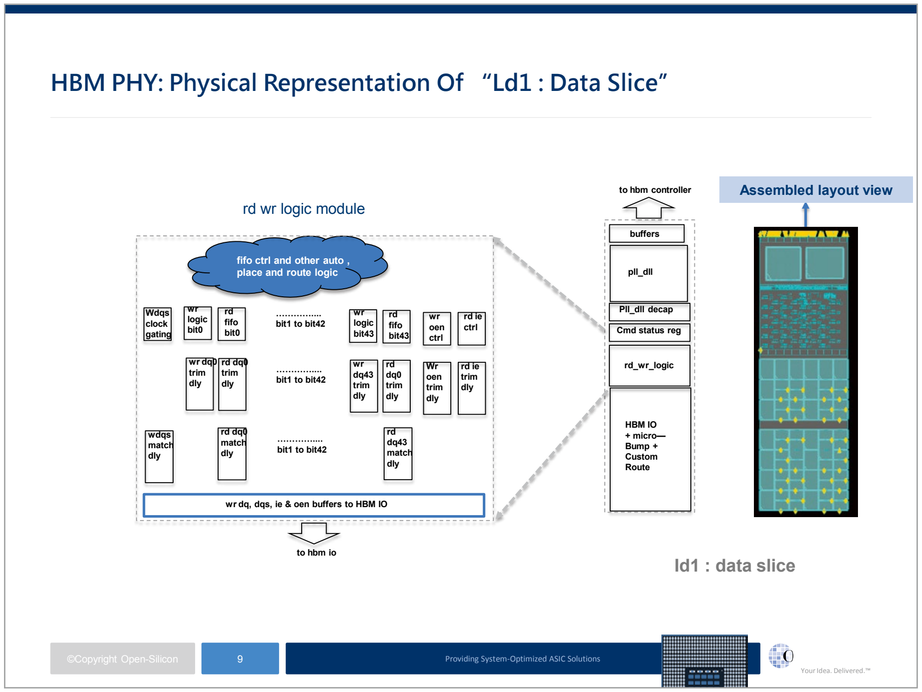
Physical View Of 2 Channel HBM PHY

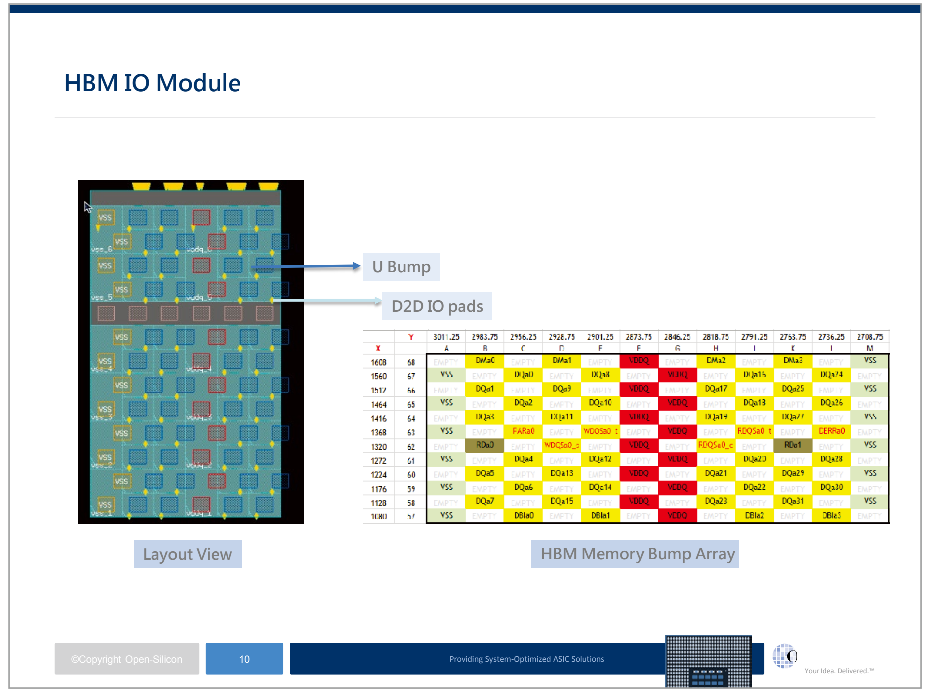


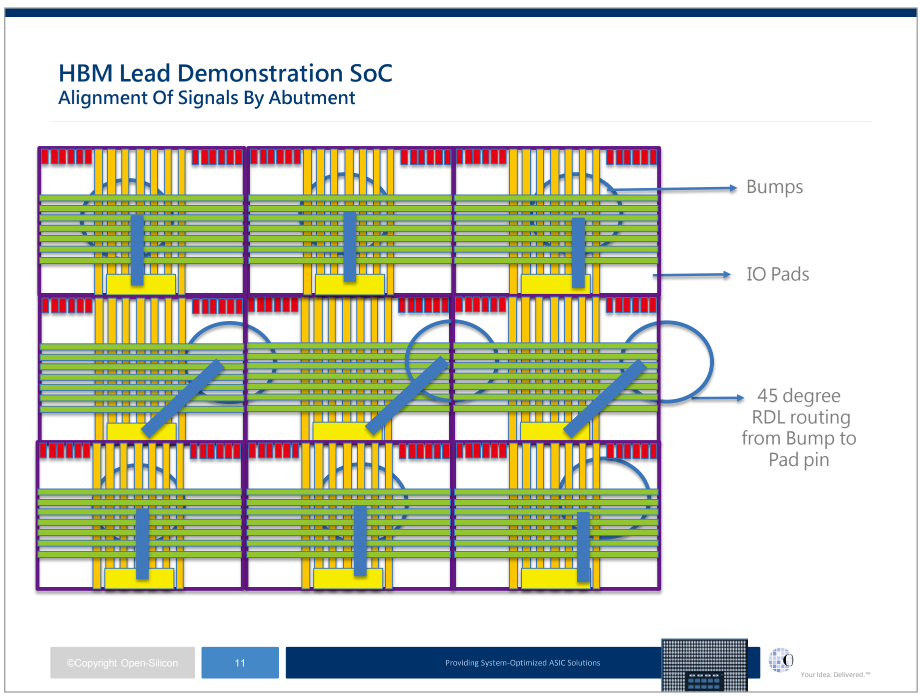
EDA Track
IP Track
EDA / IP / Services Track
Print-Only Papers

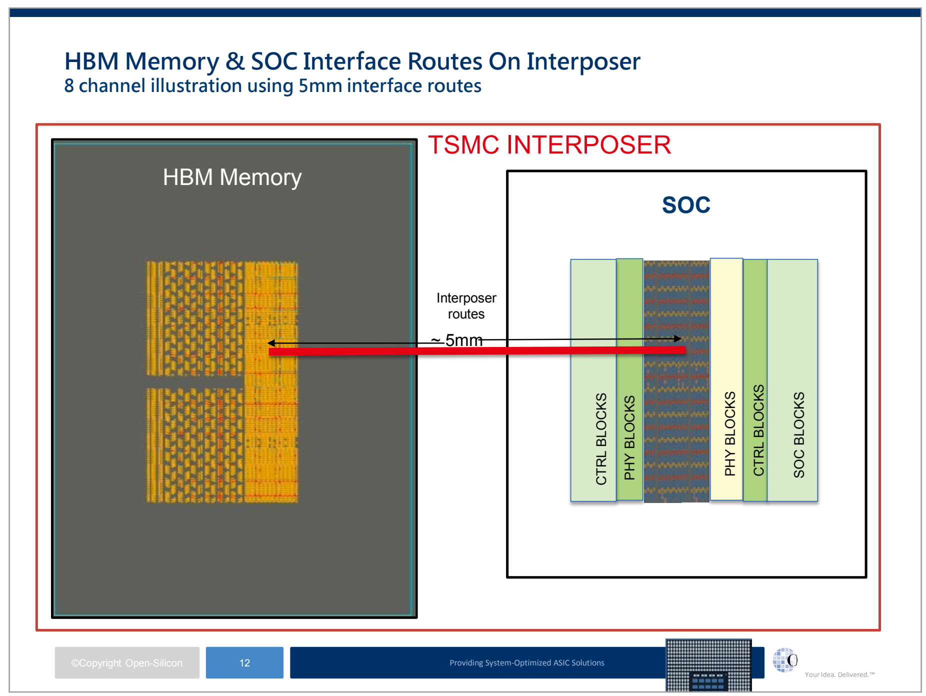
HBM IP Subsystem Implementation: 2.5D ASIC

Open-Silicon

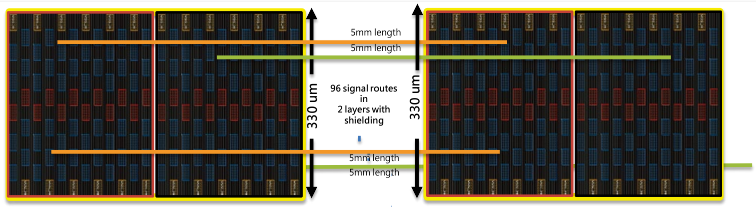








Interposer
Single Data Slice Routing



HBM – Memory data slice bump out showing 96 signal bumps

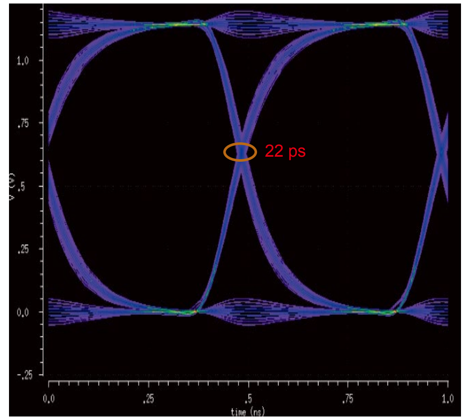
SOC– Data slice bump out showing 96 signal bumps

To effectively shield the signals from any crosstalk, all wires have a ground wire of 0.5 um width, placed at 2.1 um space from signal wire

This leaves a width of 2.1 um per signal wire

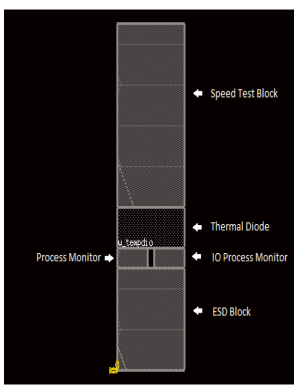
Physical Parameter	Value
Width of signal wire	2.1 um
Length of signal wire	5 mm
Thickness of signal wire	0.85 um
Spacing b/w signal & ground wire	2.1um

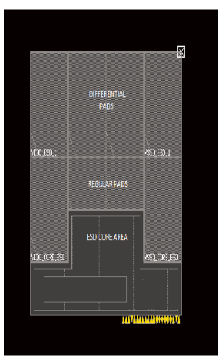
Results From Interposer Spice Simulations



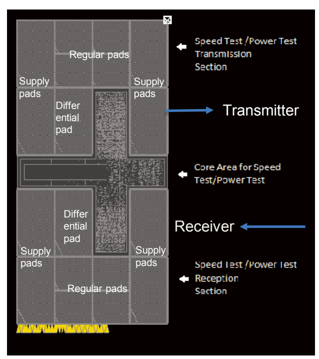
Parameter	Value
RC Delay Model	5 stage pi-network of RC
Driver Model (assumptions)	HBM2 IBIS model as Tx with 18 mA drive operating at minimum corner (125C, 1.14V), 73 ps slew (30% - 70%) at input of the IBIS IO buffer.
Receiver Model	HBM2 IBIS model as Rx operating at maximum corner (for max input cap)
Slew (30% -70%)	90.6 ps
Impact of crosstalk on data window (single corner)	13 ps (no length skew/ocv/input slew or output load variations considered)
Worst Receiver Peak-Peak	1.15 V

HBM PHY Characterization Block





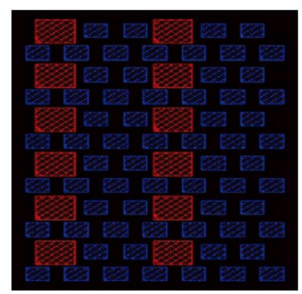
ESD test

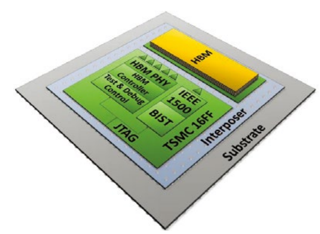


Speed test

HBM IP Physical Design for DFT
uBump and Probing

12 Rows and 6 Columns of Micro uBumps Representing a Slice





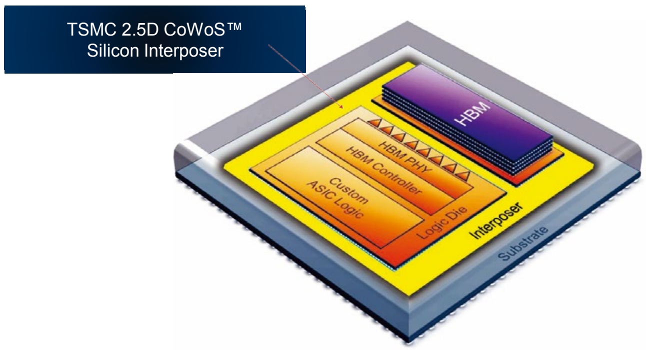
uBumps

Probe Pads

Testability feature deployed through probing mechanism for debug of ASIC SiPs with HBM memory embedded on a 2.5D interposer die

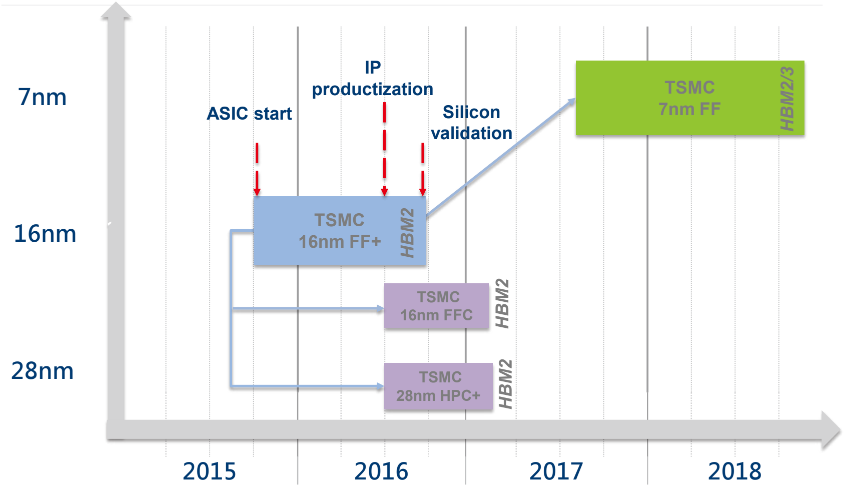
HBM IP Silicon Validation Plans With Lead Demonstration SoC

This Lead SoC Will Demonstrate
IP Subsystem ⇔ HBM Die-Stack Interoperability

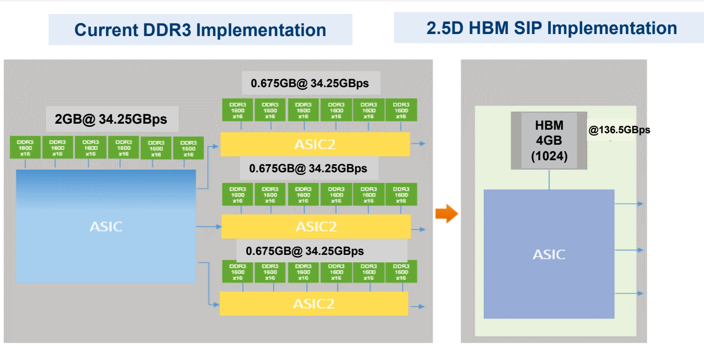


HBM ecosystem is ready for power and form-factor constrained systems in high-performance computing, networking, high-end consumer and graphics applications

HBM IP Subsystem Roadmap
TSMC Foundry



System Integration: 2.5D HBM ASIC SiP Solution
Lead Customer

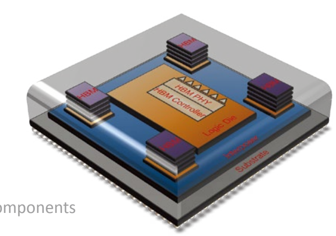


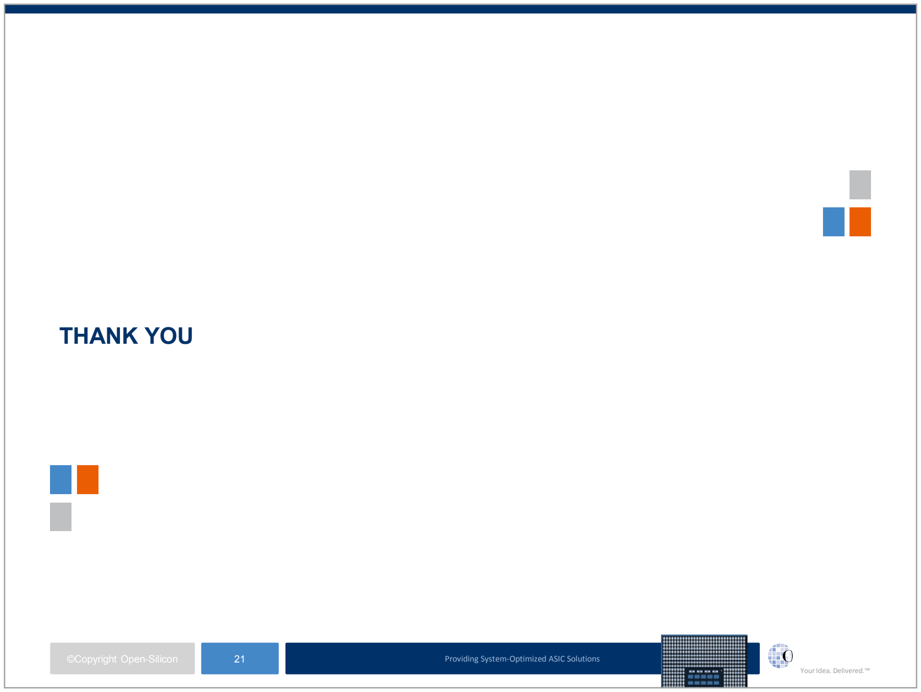
HBM Delivers Small From Factor

	Max.Bandwidth	Power Efficiency
DDR3 (x16)	4 GigaBytes per second	1.00 mW/Gigabits per second/pin
HBM2 (x1024)	256 GigaBytes per second	0.33 mW/Gigabits per second/pin

Summary

- **Full Subsystem IP Solution**
  - HBM Controller + PHY+ IO IP from one source
  - HBM IP subsystem from a single vendor minimizes integration risk
- **Performance**
  - 2Gbps per pin data-rate at longer trace lengths with custom I/O
  - Bandwidth up-to 256GB/s
- **Proven 2.5D ASIC Design and Packaging Experience**
  - Three 2.5D ASIC SiP Designs
  - HBM interface requires a specialized interposer die-to-die IO
- **Testability**
  - Probe pads allow reduced total cost of ownership
  - Loop back allows for issue-isolation for different IP subsystem components
- **Interoperability**
  - HBM 2.5D ASIC Design in progress for lead customer
  - Lead HBM SoC in TSMC 16nm CoWoS™ will demonstrate: IP Subsystem ⇔ HBM die-stack interoperability
- **Ecosystem**
  - HBM ecosystem is ready to enable supply of HBM2 ASIC SiPs




THANK YOU
©Copyright Open-Silicon
21
Providing System-Optimized ASIC Solutions
Your Idea. Delivered.™